

THE PRIMES ARE NOT METRIC POISSONIAN

ALED WALKER

ABSTRACT. It has been known since Vinogradov that, for irrational α , the sequence of fractional parts $\{\alpha p\}$ is equidistributed in \mathbb{R}/\mathbb{Z} as p ranges over primes. There is a natural second-order equidistribution property, a pair correlation of such fractional parts, which has recently received renewed interest, in particular regarding its relation to additive combinatorics. In this paper we show that the primes do not enjoy this stronger equidistribution property.

1. INTRODUCTION

Let $\mathcal{A} \subset \mathbb{N}$ be an infinite sequence of natural numbers, and let A_N denote the first N elements of \mathcal{A} . For $\alpha \in [0, 1]$, we consider the sequence $\alpha\mathcal{A}$, taken modulo 1. Recall that the sequence $\alpha\mathcal{A}$ is *equidistributed* in \mathbb{R}/\mathbb{Z} if for every interval $I \subset \mathbb{R}/\mathbb{Z}$ one has

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{x \in A_N} \mathbb{1}_I(\alpha x) = |I|. \quad (1)$$

For many arithmetic sequences \mathcal{A} of interest, the sequence $\alpha\mathcal{A}$ is equidistributed in \mathbb{R}/\mathbb{Z} for all irrational α . This is true for $\mathcal{A} = \mathbb{N}$ itself, or more generally the set of k^{th} powers for any $k \in \mathbb{N}$, and, most pertinently for us, the set of primes.

In this paper we will consider a strictly stronger notion of equidistribution. With notation as above, we define the pair correlation function

$$F(\mathcal{A}, \alpha, s, N) := \frac{1}{N} \sum_{\substack{x_i, x_j \in A_N \\ x_i \neq x_j}} \mathbb{1}_{[-s/N, s/N]}(\alpha(x_i - x_j)), \quad (2)$$

where both the interval $[-s/N, s/N]$ and the sequence $\alpha\mathcal{A}$ are considered modulo 1.

Informally, $F(\mathcal{A}, \alpha, s, N)$ counts the number of pairs $(\alpha x_i, \alpha x_j)$ such that the distance $\alpha x_i - \alpha x_j \bmod 1$ is approximately s times the average gap length of the sequence $\alpha A_N \bmod 1$. Analysing the behaviour of $F(\mathcal{A}, \alpha, s, N)$ for a specific α can require delicate Diophantine information about α (see [8], [12]), but one may instead settle for results which hold for almost all α .

In the setting of (1), *any* \mathcal{A} satisfies the equidistribution property for almost all α (the sharpest results in this direction are due to Baker [3]). However, in the setting of pair correlations, the situation is more subtle. We say that the sequence \mathcal{A} is ‘metric Poissonian’ if for almost all $\alpha \in [0, 1]$, and for all fixed $s > 0$, we have

$$F(\mathcal{A}, \alpha, s, N) = 2s(1 + o_{\mathcal{A}, \alpha, s}(1)) \quad (3)$$

as $N \rightarrow \infty$. Notice that if we had picked N i.i.d. random variables $(X_n)_{n \in [N]}$ uniformly distributed on \mathbb{R}/\mathbb{Z} , instead of the sequence $\alpha A_N \bmod 1$, then as N tends to infinity the equivalent pair correlation function would tend to $2s$ with high probability. Therefore (3) may be viewed as some strong indication that $\alpha\mathcal{A} \bmod 1$ exhibits the behaviour of a random sequence. The connection to the equidistribution property (1) was recently made rigorous: indeed, three simultaneous papers [1, 5, 14] recently showed that if (3)

holds, for some fixed α and for all s , then for the same α one has that $\alpha\mathcal{A}$ is equidistributed in \mathbb{R}/\mathbb{Z} .

One might expect that, for the classical sequences \mathcal{A} where $\alpha\mathcal{A}$ is equidistributed for irrational α , one could prove that these sequences \mathcal{A} are metric Poissonian. Indeed, for $k \geq 2$ and \mathcal{A} the set of k^{th} powers this was shown by Rudnick and Sarnak [11]. However, the sequence $\mathcal{A} = \mathbb{N}$ is *not* metric Poissonian. This follows from consideration of the continued fraction expansion of α , but is in fact a special case of a more general phenomenon, connected to the large *additive energy* of this particular set \mathcal{A} .

For a finite set $B \subset \mathbb{N}$ we define the additive energy $E(B)$ to be the number of quadruples $(b_1, b_2, b_3, b_4) \in B^4$ such that $b_1 + b_2 = b_3 + b_4$. If $|B| = N$, then we have the trivial bounds $N^2 \ll E(B) \leq N^3$. For $x \in \mathbb{R}$, let us write $\|x\|$ for $\min_{y \in \mathbb{Z}} |x - y|$. Then, for an additive quadruple (b_1, b_2, b_3, b_4) satisfying $b_1 + b_2 = b_3 + b_4$, obviously if $\|\alpha(b_1 - b_3)\| \leq \frac{s}{N}$ then $\|\alpha(b_2 - b_4)\| \leq \frac{s}{N}$. This is extremely different behaviour than that which would be seen if $\alpha b_1, \alpha b_2, \alpha b_3, \alpha b_4$ were genuinely i.i.d. uniform random variables on \mathbb{R}/\mathbb{Z} , and indeed we have the following result of Bourgain, which shows that all sets of nearly maximal energy fail to have the metric Poissonian property.

Theorem 1. [2, Appendix] *Suppose*

$$\limsup_{N \rightarrow \infty} \frac{E(A_N)}{N^3} > 0.$$

Then \mathcal{A} is not metric Poissonian.

It is clear that the sequence $\mathcal{A} = \mathbb{N}$ satisfies the hypotheses of this theorem, and therefore this sequence is not metric Poissonian.

Remarkably, a near-converse to this theorem has also been proved to be true.

Theorem 2. *Let $\delta > 0$ be fixed, and suppose that $E(A_N) \ll_{\delta} N^{3-\delta}$ for this fixed δ and for every N . Then \mathcal{A} is metric Poissonian.*

This theorem first appears as stated¹ in the recent work of Aistleitner, Larcher and Lewko [2]. It immediately implies the theorem of Rudnick-Sarnak on k^{th} powers, and also earlier work on lacunary sequences [13].

It is natural to wonder whether there is a tight energy threshold for this problem. Although the truth seems unlikely to be so clean, it is certainly interesting to consider the behaviour of specific sets \mathcal{A} which satisfy $N^{3-\varepsilon} \ll_{\varepsilon} E(A_N) \ll o(N^3)$ for all $\varepsilon > 0$. In this paper, we prove the following result (answering a question posed by Nair²).

Theorem 3 (Main Theorem). *The primes are not metric Poissonian.*

When \mathcal{A} is the set of primes one has $E(A_N) \asymp N^3(\log N)^{-1}$, so certainly the primes are not included in the range of applicability of either Theorem 1 or Theorem 2.

In [2], Bourgain constructs a sequence \mathcal{A} which is not metric Poissonian but nonetheless has $E(A_N) = o(N^3)$, thereby showing that the converse to Theorem 1 is false. A quantitative analysis of his argument shows that $E(A_N) \ll_{\varepsilon} N^3(\log \log N)^{-\frac{1}{4}+\varepsilon}$ is achievable, for any $\varepsilon > 0$. So, as an immediate corollary to Theorem 3, we have an improved bound for the smallest energy $E(A_N)$ of the initial segments of a set \mathcal{A} which is not metric Poissonian.

¹The same result may be deduced from Theorem 3.2 of Harman's earlier book [7], combined with the relevant modification of the variance estimate from page 69 of the same volume.

²at the ELAZ 2016 conference in Strobl.

For further extensions and refinements of Theorems 1 and 2, we refer the interested reader to two very recent pre-prints, [10] and [4].

Acknowledgements

The author would like to thank Prof. R. Nair for making him aware of the central question of this paper, Prof. C. Aistleitner for comments on an earlier version, and Prof. B. J. Green for his doctoral supervision. A helpful conversation was also had with Sam Chow. The work was completed while the author was a Program Associate at the Mathematical Sciences Research Institute in Berkeley, which provided excellent working conditions. The author is supported by EPSRC grant no. EP/M50659X/1.

2. PROOF OF THEOREM 3

The plan of the proof is as follows. For each fixed α , we will try to find infinitely many n such that $\|\alpha n\|$ is extremely small. Using such an n we will be able to construct a scale N and a small constant s such that $\|\alpha mn\| \leq s/N$ for some initial segment of integers m . By a variant of a well-known result concerning the exceptional set for the Goldbach problem, we may show that many such mn are represented many times as $p_i - p_j$ for two primes $p_i, p_j \leq p_N$, the N^{th} prime. Combining all these observations will enable us to conclude, provided s is small enough, that $F(\mathcal{P}, \alpha, s, N) \geq c$ for some constant $c > 2s$. Since this holds for infinitely many N , we cannot have $F(\mathcal{P}, \alpha, s, N) = 2s(1 + o_{\alpha,s}(1))$ for all almost all α and for all $s > 0$. In fact, we will show that, for almost all α , this asymptotic *fails* to hold.

We now begin to consider the details of this argument. We will use the following result of Harman on Diophantine approximation ([7, Theorem 4.2], [6]).

Theorem 4. *Let $\psi(n)$ be a non-increasing function with $0 < \psi(n) \leq \frac{1}{2}$. Suppose that*

$$\sum_n \psi(n) = \infty.$$

Let \mathcal{B} be an infinite set of integers, and let $S(\mathcal{B}, \alpha, N)$ denote the number of $n \leq N$, $n \in \mathcal{B}$, such that $\|\alpha n\| < \psi(n)$. Then for almost all α we have

$$S(\mathcal{B}, \alpha, N) = 2\Psi(N, \mathcal{B}) + O_\varepsilon(\Psi(N)^{\frac{1}{2}}(\log \Psi(N))^{2+\varepsilon}) \quad (4)$$

for all $\varepsilon > 0$, with implied constant uniform in α , where

$$\Psi(N) = \sum_{n \leq N} \psi(n)$$

and

$$\Psi(N, \mathcal{B}) = \sum_{\substack{n \leq N \\ n \in \mathcal{B}}} \psi(n).$$

This theorem may be thought of as a flexible version of Khintchine's theorem on Diophantine approximation, in which one can further pass to approximations coming from a set \mathcal{B} , provided \mathcal{B} is relatively dense. The quality of the error term in this theorem is much better than we need in our application, although it is important that there is no dependence on N except through $\Psi(N)$. Earlier results of this type include a $\log N$ factor in the error, which would not have been adequate.

The other technical tool will be the standard bound on the size of the exceptional set in a Goldbach-like problem.

Theorem 5. *For a large quantity X , and natural number $n \leq X$, define*

$$r(n) := \sum_{\substack{p_i, p_j \leq X \\ p_i - p_j = n}} \log p_i \log p_j.$$

Then for any $B > 0$, and for all but $O_B(\frac{X}{\log^B X})$ exceptional values of $n \leq X$, we have the approximation

$$r(n) = \mathfrak{S}(n)J(n) + O_B\left(\frac{X}{\log^B X}\right), \quad (5)$$

where

$$\mathfrak{S}(n) := \begin{cases} 2 \prod_{p \geq 3} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{\substack{p|n \\ p \geq 3}} \frac{p-1}{p-2} & n \text{ even,} \\ 0 & n \text{ odd} \end{cases}$$

is the singular series, and

$$J(n) = \int_{-\infty}^{\infty} \left| \int_0^X e(\beta u) du \right|^2 e(-\beta n) d\beta$$

is the singular integral.

Proof. This result follows by trivial modifications of the usual argument for the binary Goldbach problem, originally due (independently) to van der Corput, Estermann, and Čudakov. The clearest reference is [15, Chapter 3.2], or, for a more modern approach, one may consider the proof of Theorem 19.1 in [9, Chapter 19]. \square

We combine these two key ingredients in the following proposition.

Proposition 6. *There exists a small absolute $c > 0$, such that for almost all $\alpha \in [0, 1]$, and for all fixed $s > 0$, there exist infinitely many n satisfying:*

(i) $\|\alpha n\| < \frac{s}{n \log n}$

(ii) *At least $c \log n$ of the numbers $n, 2n, \dots, \lfloor \frac{1}{10} \log n \rfloor n$ are expressible in at least $c \frac{n}{\log n}$ ways as the difference $p_1 - p_2$ of two primes $p_1, p_2 \leq \frac{1}{2} n \log n$.*

Proof of Proposition 6. Let $c > 0$ be a quantity to be specified later. With this c , let \mathcal{B} be the set of natural numbers n which satisfy (ii), and let $\psi(n) = \min(\frac{1}{2}, \frac{s}{n \log n})$. It is to this \mathcal{B} and this ψ that we will apply Theorem 4.

We claim that \mathcal{B} is relatively dense. Indeed, let K be a large integer, and let n and m be natural numbers restricted to the ranges $K \leq n < 2K$ and $1 \leq m \leq \lfloor \frac{1}{10} \log 2K \rfloor$. For notational convenience we let X denote the quantity $\frac{1}{2} K \log K$, and we consider Theorem 5 with this X . We say that the pair (n, m) is *exceptional* if nm lies in the exceptional set from Theorem 5 for which the asymptotic formula (5) fails to hold. [Note that $nm \leq X$, so Theorem 5 applies in this setting.]

The map $(n, m) \mapsto nm$ is at most $\frac{1}{10} \log 2K$ -to-1, due to the restricted range of m . Since the exceptional set from Theorem 5 has size at most $O_B(\frac{K}{\log^B K})$, for all $B > 0$, we conclude that there are at most $O_B(\frac{K}{\log^B K})$ exceptional pairs (n, m) , for all $B > 0$. So certainly there are at least $(1 - O_B(\log^{-B} K))K$ values of $n \in [K, 2K)$ such that the asymptotic formula for $r(nm)$ holds for all $m \leq \frac{1}{10} \log 2K$. Let D_K denote this set of n .

D_K is certainly very dense in $[K, 2K)$, and we claim further that $D_K \subset \mathcal{B}$, provided we choose c small enough. Combining the different scales K will allow us to show that \mathcal{B} is suitably dense. Indeed, let us analyse the asymptotic formula for $r(nm)$. When

nm is even, the singular series $\mathfrak{S}(nm)$ is always $\Omega(1)$. By Fourier inversion, the singular integral is exactly $(\mathbb{1}_{[0,X]} * \mathbb{1}_{[-X,0]})(nm)$, which is $\Omega(X)$ provided that $|nm| \leq \frac{1}{5}X$. But, by the choice of ranges for n and m , this inequality is always satisfied. So, for all $n \in D_K$ and for all $m \leq \frac{1}{10} \log n$, such that nm is even, we have $r(nm) \gg X$.

Removing the log weights on the primes, and recalling the definition of X , in particular we notice that there is some small absolute constant c_1 such the following holds: if $n \in D_K$ and if nm is even, there are at least $c_1 K / \log K$ pairs of primes (p_i, p_j) with $p_i, p_j \leq \frac{1}{2} K \log K$ such that $p_i - p_j = nm$. By the definition of \mathcal{B} , provided c is chosen smaller than $\min(c_1, \frac{1}{100})$, we have that $D_K \subset \mathcal{B}$.

We may now prove that $\Psi(\mathcal{B}, N) \gg_s \Psi(N)$ for large N . Indeed,

$$\begin{aligned} \Psi(\mathcal{B}, N) &= \sum_{\substack{n \leq N \\ n \in \mathcal{B}}} \min\left(\frac{1}{2}, \frac{s}{n \log n}\right) \\ &\gg -O_s(1) + \sum_{k=k_0}^{\lfloor \log_2 N \rfloor - 1} \sum_{\substack{2^k \leq n < 2^{k+1} \\ n \in \mathcal{B}}} \frac{s}{2^k \log(2^k)} \\ &\gg -O_s(1) + \sum_{k=k_0}^{\lfloor \log_2 N \rfloor - 1} \sum_{2^k \leq n < 2^{k+1}} (1 - O_B(k^{-B})) \frac{s}{k 2^k} \\ &\gg -O_s(1) + \sum_{k=1}^{\lfloor \log N \rfloor} \frac{s}{k} \\ &\gg_s \log \log N \\ &\gg_s \Psi(N). \end{aligned}$$

Therefore, applying Theorem 4 to this set \mathcal{B} and this function ψ , the main term from the conclusion of Theorem 4 dominates the error term, and we conclude that for almost all α there are infinitely many $n \in \mathcal{B}$ satisfying $\|\alpha n\| < \frac{s}{n \log n}$. The proposition is proved. \square

With this moderately technical proposition proved, the deduction of Theorem 3 is extremely short.

Proof of Theorem 3. Let $\Omega \subset [0, 1]$ be the full-measure set of α for which Proposition 6 holds. Let c be the constant from Proposition 6, and fix some s satisfying $0 < 2s < c^2$. Let $\alpha \in \Omega$, and fix a large N to be one of the infinitely many natural numbers which satisfy the conclusions of Proposition 6.

By construction, we know that

$$\|\alpha N\| < \frac{s}{N \log N}.$$

Therefore, for all $d \leq \frac{1}{10} \log N$, we have

$$\|\alpha d N\| < \frac{s}{N}.$$

But by the second conclusion of Proposition 6, this implies that there are at least $c^2 N$ pairs of distinct primes $p_i, p_j \leq \frac{1}{2} N \log N$ such that

$$\|\alpha(p_i - p_j)\| < \frac{s}{N}.$$

Since $p_N \sim N \log N$, and N is large, this certainly implies that

$$F(\mathcal{P}, \alpha, s, N) \geq c^2 > 2s.$$

This holds for infinitely many N , and therefore for all $\alpha \in \Omega$ we have

$$F(\mathcal{P}, \alpha, 2s, N) \neq 2s(1 + o(1))$$

as $N \rightarrow \infty$. Since Ω has measure 1, Theorem 3 is proved. \square

REFERENCES

- [1] Christoph Aistleitner, Thomas Lachmann, and Florian Pausinger. *Pair correlations and equidistribution*. preprint at <https://arxiv.org/abs/1612.05495>.
- [2] Christoph Aistleitner, Gerhard Larcher, Mark Lewko, and with an appendix by Jean Bourgain. *Additive Energy and the Hausdorff dimension of the exceptional set in metric pair correlation problems*. To appear in *Israel Journal of Mathematics*, preprint at <https://arxiv.org/abs/1606.03591>.
- [3] R. C. Baker. Metric number theory and the large sieve. *J. London Math. Soc. (2)*, 24(1):34–40, 1981.
- [4] T. Bloom, S. Chow, A. Gafni, and A. Walker. Additive energy and the metric poissonian property. preprint at <https://arxiv.org/abs/1709.02634>.
- [5] Sigrid Grepstad and Gerhard Larcher. *On pair correlation and discrepancy*. preprint at <https://arxiv.org/abs/1612.08008>.
- [6] Glyn Harman. Metric Diophantine approximation with two restricted variables. I. Two square-free integers, or integers in arithmetic progressions. *Math. Proc. Cambridge Philos. Soc.*, 103(2):197–206, 1988.
- [7] Glyn Harman. *Metric number theory*, volume 18 of *London Mathematical Society Monographs. New Series*. The Clarendon Press, Oxford University Press, New York, 1998.
- [8] D. R. Heath-Brown. Pair correlation for fractional parts of αn^2 . *Math. Proc. Cambridge Philos. Soc.*, 148(3):385–407, 2010.
- [9] Henryk Iwaniec and Emmanuel Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004.
- [10] T. Lachmann and N. Technau. On exceptional sets in the metric poissonian pair correlations problem. preprint at <https://arxiv.org/abs/1708.08599>.
- [11] Zeév Rudnick and Peter Sarnak. The pair correlation function of fractional parts of polynomials. *Comm. Math. Phys.*, 194(1):61–70, 1998.
- [12] Zeév Rudnick, Peter Sarnak, and Alexandru Zaharescu. The distribution of spacings between the fractional parts of $n^2\alpha$. *Invent. Math.*, 145(1):37–57, 2001.
- [13] Zeév Rudnick and Alexandru Zaharescu. A metric result on the pair correlation of fractional parts of sequences. *Acta Arith.*, 89(3):283–293, 1999.
- [14] Stefan Steinerberger. *Localized Quantitative Criteria for Equidistribution*. preprint at <https://arxiv.org/abs/1701.08323>.
- [15] R. C. Vaughan. *The Hardy-Littlewood method*, volume 125 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, second edition, 1997.

ALED WALKER, MATHEMATICAL INSTITUTE, UNIVERSITY OF OXFORD, ANDREW WILES BUILDING, RADCLIFFE OBSERVATORY QUARTER, WOODSTOCK ROAD, OXFORD, OX2 6GG.

walker@maths.ox.ac.uk